STAT 512 hw 6

1. Let $Y_1, \dots, Y_n$ be a random sample from the distribution with pdf given by

$$f_Y(y) = \frac{\Gamma(2+\beta)}{\Gamma(\beta)} y(1-y)^{\beta-1}, \quad 0 < y < 1.$$

(a) Find $\mathbb{E}\bar{Y}_n$. *Hint: Try to recognize the pdf.*

We recognize the pdf as that of the $\text{Beta}(2, \beta)$ distribution, which has expected value equal to $2/(2+\beta)$. Since the sample mean has expected value equal to the population mean, we have $\mathbb{E}\bar{Y}_n = 2/(2+\beta)$.

(b) Find $\text{Var}\,\bar{Y}_n$.

The variance of the sample mean is equal to the population mean divided by the sample size n, so we have

$$\text{Var}\,\bar{Y}_n = \frac{2\beta}{n(2+\beta)^2(2+\beta+1)}.$$

(c) Show that $\bar{Y}_n$ is a consistent estimator of $2/(2+\beta)$.

Note that $2/(2+\beta)$ is the population mean.

If $\text{Bias}\,\bar{Y}_n$ and $\text{Var}\,\bar{Y}_n$ converge to 0 as $n \to \infty$, then $\bar{Y}_n$ is a consistent estimator. This is the case, since $\text{Bias}\,\bar{Y}_n = 0$ and $\text{Var}\,\bar{Y}_n$ has n in its denominator, so that it will converge to 0 as $n \to \infty$.

We could also argue from the WLLN: Since the variance of the population, which is $2\beta/[(2+\beta)^2(2+\beta+1)]$, is finite, the sample mean converges in probability to the population mean, i.e. the sample mean is a consistent estimator of the population mean.

(d) Propose a function of $\bar{Y}_n$ which is a consistent estimator of β.

We have

$$\bar{Y}_n \xrightarrow{\text{p}} \frac{2}{2+\beta} \implies \frac{2-2\bar{Y}_n}{\bar{Y}_n} \xrightarrow{\text{p}} \beta,$$

so that

$$\hat{\beta} = \frac{2-2\bar{Y}_n}{\bar{Y}_n}$$

is a consistent estimator of β. We can obtain this by setting $\bar{Y}_n = 2/(2+\beta)$ and solving for β. Then we note that the function

$$\bar{Y}_n \mapsto \frac{2-2\bar{Y}_n}{\bar{Y}_n}$$

is continuous for $\bar{Y}_n \in (0, 1)$.

2. Let $X_1, \dots, X_n$ be a random sample from the distribution with cdf given by

$$F_X(x) = \begin{cases} 1 - \left(\frac{c}{x}\right)^\alpha, & x \geq c \\ 0, & x < c \end{cases}$$

(a) Find the population pdf.

Taking the derivative of the cdf, we obtain

$$f_X(x) = \alpha c^\alpha x^{-(\alpha+1)}, \quad x \geq c.$$

(b) Find the pdf of $X_{(1)}$.

We have

$$\begin{aligned} f_{X_{(1)}}(x) &= n\left[1 - \left(1 - \left(\frac{c}{x}\right)^\alpha\right)\right]^{n-1} \alpha c^\alpha x^{-(\alpha+1)} \\ &= n\alpha c^{n\alpha} x^{-(n\alpha+1)}, \quad x \geq c. \end{aligned}$$

(c) Show that $X_{(1)}$ is a consistent estimator of c by directly showing that

$$\lim_{n\to\infty} P(|X_{(1)} - c| < \varepsilon) = 1$$

for every $\varepsilon > 0$.

We have

$$\begin{aligned} P(|X_{(1)} - c| < \varepsilon) &= P(c - \varepsilon < X_{(1)} < c + \varepsilon) \\ &= P(c < X_{(1)} < c + \varepsilon) \qquad \text{(support begins at } c\text{)} \\ &= \int_c^{c+\varepsilon} n\alpha c^{n\alpha} x^{-(n\alpha+1)} dx \\ &= c^{n\alpha}\left[c^{-n\alpha} + (c+\varepsilon)^{-n\alpha}\right] \\ &= 1 - \left(\frac{c}{c+\varepsilon}\right)^{n\alpha} \end{aligned}$$

Since $\varepsilon > 0$, the second term goes to zero as $n \to \infty$, so that the limit is 1.

3. Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$.

(a) For $n = 10$, give the value of a such that $P\left(\bar{X}_n - aS_n/\sqrt{n} < \mu < \bar{X}_n + aS_n/\sqrt{n}\right) = 0.99$.

The value of a needs to be the upper 0.005 quantile of the t_9 distribution, which is `qt(.995,9)` $= 3.249836$.

(b) For $n = 10$, find $P\left(\bar{X}_n > \mu + 2S_n/\sqrt{n}\right)$.

We have

$$P\left(\bar{X}_n > \mu + 2S_n/\sqrt{n}\right) = P\left(\sqrt{n}(\bar{X}_n - \mu)/S_n > 2\right) = P(T > 2), \quad T \sim t_9.$$

So the answer is `1-pt(2,9)` $= 0.03827641$.

(c) For $n = 10$, find $P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right)$.

$$\begin{aligned} P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right) &= P\left(-2 < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < 2\right) \\ &= P(-2 < T < 2), \quad T \sim t_9. \end{aligned}$$

So the answer is `pt(2,9) - pt(-2,9)` $= 0.9234472$.

(d) Give $\lim_{n\to\infty} P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right)$.

We have

$$\begin{aligned} \lim_{n\to\infty} P\left(\bar{X}_n - 2S_n/\sqrt{n} < \mu < \bar{X}_n + 2S_n/\sqrt{n}\right) &= \lim_{n\to\infty} P\left(-2 < \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} < 2\right) \\ &= P(-2 < Z < 2), \quad Z \sim \text{Normal}(0,1). \end{aligned}$$

So the answer is `pnorm(2) - pnorm(-2)` $= 0.9544997$.

4. Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Normal}(\mu, \sigma^2)$.

(a) For $n = 10$, give values of a and b such that $P\left(aS_n^2 < \sigma^2 < bS_n^2\right) = 0.90$.

Writing

$$\begin{aligned} 0.90 = P\left(aS_n^2 < \sigma^2 < bS_n^2\right) &= P\left(\frac{1}{aS_n^2} > \frac{1}{\sigma^2} > \frac{1}{bS_n^2}\right) \\ &= P\left(\frac{n-1}{a} > \frac{(n-1)S_n^2}{\sigma^2} > \frac{n-1}{b}\right), \end{aligned}$$

we see that we can choose a and b which satisfy

$$\frac{n-1}{a} = \chi^2_{n-1,0.05} \iff a = \frac{n-1}{\chi^2_{n-1,0.05}} \quad \text{and} \quad \frac{n-1}{b} = \chi^2_{n-1,1-0.05} \iff b = \frac{n-1}{\chi^2_{n-1,1-0.05}}.$$

For $n = 10$ we have

$$a = \texttt{9/qchisq(.95,9)} = 0.531947 \quad \text{and } b = \texttt{9/qchisq(.05,9)} = 2.706675.$$

(b) For $n = 10$, find $P\left((0.473)S_n^2 < \sigma^2 < (3.333)S_n^2\right)$.

We have

$$P\left((0.473)S_n^2 < \sigma^2 < (3.333)S_n^2\right) = P\left(\frac{n-1}{0.473} > \frac{(n-1)S_n^2}{\sigma^2} > \frac{n-1}{3.333}\right).$$

For $n = 10$ this becomes

$$P\left(\frac{9}{0.473} > \frac{9S_9^2}{\sigma^2} < \frac{9}{3.333}\right) = P\left(\frac{9}{0.473} > W > \frac{9}{3.333}\right), \quad W \sim \chi_9^2.$$

So the answer is `pchisq(9/0.473,9) - pchisq(9/3.333,9)` $= 0.9500436$.

(c) Give $\lim_{n\to\infty} P(-0.01 < S_n^2 - \sigma^2 < 0.01)$.

Since under these settings S_n^2 is a consistent estimator of σ^2 (shown in the notes), the limit is 1.

5. Suppose $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\lambda)$. Then according to the central limit theorem

$$\frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}} \to^D Z,$$

as $n \to \infty$, where $Z \sim \text{Normal}(0, 1)$.

(a) Use the above result to find

$$\lim_{n\to\infty} P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right).$$

Since the mean and variance of the Exponential(λ) distribution are λ and λ^2, respectively, we have

$$\lim_{n\to\infty} P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right) = P(1 < Z), \quad Z \sim \text{Normal}(0, 1),$$

so the answer is `1 - pnorm(1)` $= 0.1586553$.

(b) Use mgfs to show that $\sqrt{n}\bar{X}_n/\lambda \sim \text{Gamma}(n, 1/\sqrt{n})$.

We have

$$\begin{aligned} M_{\sqrt{n}\bar{X}_n/\lambda}(t) &= M_{\bar{X}_n}(\sqrt{n}t/\lambda) \\ &= M_{(X_1+\cdots+X_n)/n}(\sqrt{n}t/\lambda) \\ &= M_{X_1+\cdots+X_n}(t/(\sqrt{n}\lambda)) \\ &= [M_{X_1}(t/(\sqrt{n}\lambda))]^n \\ &= [(1-\lambda(t/(\sqrt{n}\lambda)))^{-1}]^n \\ &= (1-(1/\sqrt{n})t)^{-n}, \end{aligned}$$

which we recognize as the mgf of the $\text{Gamma}(n, 1/\sqrt{n})$ distribution.

(c) Use the previous result to compute the probability

$$P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right)$$

for $\lambda = 10$ and $n = 10, 20, 30, 100, 500, 1000, 10000$. The numbers should approach your answer from part (a). *Hint: Use the* `pgamma` *function.*

We have

$$\begin{aligned} P\left(1 < \frac{\bar{X}_n - \lambda}{\lambda/\sqrt{n}}\right) &= P\left(1 + 1/\sqrt{n} < \sqrt{n}\bar{X}_n/\lambda\right) \\ &= P(1+\sqrt{n} < G), \quad G \sim \text{Gamma}(n, 1/\sqrt{n}). \end{aligned}$$

The following R code computes the probabilities:

```
lambda <- 10
n <- c(10,20,30,100,500,1000,10000)
1 - pgamma(1 + sqrt(n),shape = n,scale=1/sqrt(n))
```

The probabilities are

```
0.1553584 0.1569141 0.1574648 0.1582787 0.1585771 0.1586158 0.1586512
```

We note that these approach the value of $1 - \Phi(1) = 0.1586553$, where Φ is the standard Normal cdf.

(d) Suppose we observe $\bar{X}_n = 20.2$ on a sample of size $n = 50$. Give a 95% CI for λ based on

i. the exact pivot quantity result $\sqrt{n}\bar{X}_n/\lambda \sim \text{Gamma}(n, 1/\sqrt{n})$.

To construct a $(1-\alpha)100\%$ confidence interval for λ based on the pivot quantity

$\sqrt{n}\bar{X}_n/\lambda$, we write

$$P\left(G_{n,1/\sqrt{n},1-\alpha/2} < \frac{\sqrt{n}\bar{X}_n}{\lambda} < G_{n,1/\sqrt{n},\alpha/2}\right) = 1-\alpha,$$

where $G_{n,1/\sqrt{n},\xi}$ is the upper ξ quantile of the $\text{Gamma}(n, 1/\sqrt{n})$ distribution for $\xi \in (0,1)$. Rearranging this until λ appears alone in the middle results in

$$P\left(\frac{\sqrt{n}\bar{X}_n}{G_{n,1/\sqrt{n},\alpha/2}} < \lambda < \frac{\sqrt{n}\bar{X}_n}{G_{n,1/\sqrt{n},1-\alpha/2}}\right) = 1-\alpha.$$

For $\alpha = 0.05$, $n = 50$ and $\bar{X}_n = 20.2$, the confidence interval is given by

$$\left(\frac{\sqrt{50}(20.2)}{9.16136}, \frac{\sqrt{50}(20.2)}{5.248283}\right) = (15.59109, 27.21568),$$

where

$$9.16136 = \texttt{qgamma(.975,shape=n,scale=1/sqrt(n))}$$
$$5.248283 = \texttt{qgamma(.025,shape=n,scale=1/sqrt(n))}.$$

ii. the asymptotic pivot quantity result $\sqrt{n}(\bar{X}_n - \lambda)/\lambda \xrightarrow{\text{D}} Z$, $Z \sim \text{Normal}(0,1)$.

Because of the central limit theorem we can write

$$\lim_{n\to\infty} P\left(-z_{\alpha/2} < \sqrt{n}(\bar{X}_n - \lambda)/\lambda < z_{\alpha/2}\right) = 1-\alpha,$$

which we can by rearrangement write as

$$\lim_{n\to\infty} P\left(\frac{\sqrt{n}\bar{X}_n}{\sqrt{n} - z_{\alpha/2}} < \lambda < \frac{\sqrt{n}\bar{X}_n}{\sqrt{n} + z_{\alpha/2}}\right) = 1-\alpha.$$

So an approximate 95% confidence interval for λ based on $n = 50$ and $\bar{X}_n = 20.2$ is given by

$$(15.81609, 27.94613).$$

It is quite similar to the exact one!

Hint: Use `qgamma(.,shape = ., scale = .)` *to obtain quantiles of the Gamma distribution.*

6. Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$, and let $\hat{p}_n = n^{-1}(X_1 + \cdots + X_n)$.

   (a) Give an exact expression for $P(1/2 < \hat{p}_n)$.

We have

$$
\begin{aligned}
P(1/2 < \hat{p}_n) &= P(1/2 < Y/n), \quad Y \sim \text{Binomial}(n,p) \\
&= P(n/2 < Y) \\
&= \sum_{y=\lceil n/2 \rceil}^{n} \binom{n}{y} p^y (1-p)^{n-y}.
\end{aligned}
$$

(b) Evaluate your expression from part (a) for $n = 200$ and $p = 4/9$.

We can write

$$P(1/2 < \hat{p}_n) = 1 - \sum_{y=0}^{\lfloor n/2 \rfloor} \binom{n}{y} p^y (1-p)^{n-y}$$

and evaluate it with $n = 200$ and $p = 4/9$ with

`1 - pbinom(100,200,4/9) = 0.04957838`.

(c) Find an approximation to $P(1/2 < \hat{p}_n)$ when $n = 200$ and $p = 4/9$ using the fact that

$$\frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \to^D Z,$$

as $n \to \infty$, where $Z \sim \text{Normal}(0,1)$.

We have

$$
\begin{aligned}
P(1/2 < \hat{p}_n) &= P\left(\frac{1/2 - 4/9}{\sqrt{4/9(1-4/9)/200}} < \frac{\hat{p}_n - 4/9}{\sqrt{4/9(1-4/9)/200}}\right) \\
&\to P\left(1.581139 < Z\right), \quad Z \sim \text{Normal}(0,1).
\end{aligned}
$$

as $n \to \infty$. And $P\left(1.581139 < Z\right) =$ `1 - pnorm(1.581139)` $= 0.05692313$.

(d) Suppose that for $n = 200$, $\hat{p} = 0.64$. Give an approximate 95% CI for p based on this data.

According to the formula in the notes, an approximate 95% confidence interval for p would be

$$0.64 \pm 1.959964\sqrt{0.64(1-0.64)/200} = (0.5069532, 0.7730468).$$

7. Consider drawing a random sample $X_1, \ldots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda)$ and computing the interval

$$\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}.$$

(a) Give $\lim_{n\to\infty} P\left(\bar{X}_n - S_n/\sqrt{n} < \lambda < \bar{X}_n + S_n/\sqrt{n}\right)$.

We have

$$\lim_{n\to\infty} P\left(\bar{X}_n - S_n/\sqrt{n} < \lambda < \bar{X}_n + S_n/\sqrt{n}\right) = \lim_{n\to\infty} P\left(-1 < \frac{\bar{X}_n - \lambda}{S_n/\sqrt{n}} < 1\right) = P(-1 < Z < 1),$$

where $Z \sim \text{Normal}(0,1)$, so the answer is `pnorm(1) - pnorm(-1)` $= 0.6826895$.

(b) For small values of n, the interval $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ will *not* contain λ with the desired probability of $1-\alpha$. For large n, however, by the central limit theorem and an application of Slutzky's theorem, the probability that $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ contains λ should be close to $1-\alpha$.

The *coverage* of a confidence interval is the probability that it contains its target. The *nominal coverage* $1-\alpha$ is the stated and desired coverage, which may differ from the actual coverage.

Conduct some simulations to estimate the coverage of the confidence interval $\bar{X}_n \pm z_{\alpha/2} S_n/\sqrt{n}$ for $\alpha = 0.05$ when $\lambda = 20$ for the sample sizes $n = 5, 10, 15, 25, 50, 100$. For each value of n, generate 1000 realizations of the interval. Here is partial code:

```
covered <- logical(S) # vector to store TRUE/FALSE values

for(s in 1:S)
{
    # generate random sample from Exp(lambda)
    X <- rexp(n,1/lambda)

    X.bar <- mean(X)
    S.n <- sd(X)

    L <- X.bar - qnorm(1-alpha/2) * S.n / sqrt(n)
    U <- X.bar + qnorm(1-alpha/2) * S.n / sqrt(n)

    # check whether interval contained lambda
    covered[s] <- ( L < lambda ) & ( U > lambda )
}

# compute proportion of TRUE values
coverage <- mean(covered)

coverage
```

What coverages do you get for the sample sizes $n = 5, 10, 15, 25, 50, 100$? For what sample sizes do you advise using this confidence interval (for what sample sizes is the coverage close to 0.95)?

Coverages should be similar to

```
0.841 0.861 0.888 0.922 0.929 0.948
```

so they approach 95% as n increases.

Optional (do not turn in) problems for additional study from Wackerly, Mendenhall, Scheaffer, 7th Ed.:

- 8.60, 8.63
- 9.17, 9.20, 9.25 (about consistency)